Lore

SUBJECTIVE ASSESSMENT OF PILOT WORKLOAD IN THE ADVANCED FIGHTER COCKPIT

Stephen D. Detro
The BDM Corporation

ABSTRACT

Pilot workload was subjectively evaluated in a mockup of an advanced fighter cockpit. Reflective and projective SWAT rating data were collected for representative mission segments to provide a comparison of pilot workload between current fighters and an advanced fighter cockpit concept, After reflectively rating six mission segments in their baseline fighter, twelve experienced fighter pilots roleplayed these same scripted mission segments as they projected their probable workload in an advanced fighter mockup. To determine the affect that new technologies might have on the cockpit, prototype controls, displays and Chemical-Biological-Radiological (CBR) aircrew protection gear were integrated into the cockpit mockup and subjective workload ratings were collected. The cockpit mockup was designed with seven display surfaces simulated by rear projection 35mm slides, and three simulated modes of control: voice, touch, and Hands On Throttle and Stick (HOTAS). Nonparametric statistics performed on the reflective and projective SWAT ratings showed that the pilot workload was in some cases, rated significantly lower in the advanced fighter mockup than in the present day baseline fighter when CBR gear was not worn. When CBR aircrew protection was worn in the mockup, pilot workload was rated above present day fighter levels without CBR gear for three of the six mission segments. In two of the six mission segments in the mockup, workload was rated as being significantly greater than baseline when CBR protection equipment was worn, Additional information was collected from pilots in the areas of information display, control allocation and redundancy, avionics moding, operation logic, anthropometrics, and cockpit layout.

INTRODUCTION

Throughout advanced cockpit concept development, the component which seems to be the limiting factor is the human operator. While aircraft technology and performance parameters continue to improve capability, the human operator continues to perceive and function at the same rate and capacity as in the past, Advanced cockpit designs provide an unprecedented capability and necessity for information display and control interoperability, The potential for an intolerable increase in pilot workload due to information processing or systems operation overload poses a problem of serious consequence.

Historically, the evaluations of human-machine interactions in aircraft cockpits have not been performed until during the middle or late stages of system design. By this point in the process, the system definition is well fixed and changes are not easily implemented. On the other hand, identification of concept deficiencies early in the paper stage of the design or predesign process can objectively support later decisions involving critical design tradeoffs, When changing the location of a single indicator in an aircraft cockpit can cost millions even before a single production model aircraft has been built, this early validation of cockpit concepts is crucial.

The evaluation of tentative cockpit designs poses many experimental measurement problems. Since the system does not actually exist and is not operational, the analytic tools such as task analysis and time line analysis are not appropriate, Alternate measurement techniques that produce quantifiable data that can be used for predicting are required. Two procedural variants of the Subjective Workload Assessment Technique (SWAT) (Reid, Shingledecker, and Eggemeier, 1981) have been developed to help meet this need. The functional refinement of SWAT in pre-design evaluation efforts has produced both a reflective (Arbak, Shew, and Simmons, 1984) and a projective workload measurement technique (Reid, and Shingledecker, 1984). In a procedure based on the structured combination of the SWAT technique and another subjective technique called Ground Attack Tactics Survey (GATS) (Greene, Arbak, Courtright, and O'Donnell, 1981), quantitative data can be effectively derived from structured interviews with "expert operators" of similar systems. The GATS methodology, like that of the current study, calls for a structured investigation of combat mission "chokepoints" or high workload segments that have been identified to systematically reveal tasks and subtasks that contain excessively high workload. This study incorporates the use of both SWAT techniques and a cockpit mockup variation of the GATS technique,

This study was conducted at the Crew Systems Development Branch of the Flight Dynamics Laboratory, WPAFB, OH to explore the workload implications of advanced crew system control and display concepts for an advanced fighter cockpit when prototype aircrew protection equipment (Chemical, Biological, and Radiological (CBR)) was also in use. Candidate cockpit concepts were evaluated within a mission scenario which was made up of six representative advanced fighter mission segments. A cockpit mockup was built to represent control and display technologies targeted for operational use in the 1990's.

The study hypothesis was that future cockpit automation and multiple control options would decrease workload levels (compared to present day fighters), while the inclusion of CBR gear might negate some of the workload reductions made possible by the use of advanced cockpit technologies.

METHOD

Subjects

Twelve tactically experienced USAF pilots, with an average of 1,988 fighter hours, participated in this study. The pilots were volunteers selected from the pilot population at Wright-Patterson AFB, OH. Their fighter experience included the F-4, F-15, F-5, and A-7 aircraft.

Apparatus

The mockup consisted of a wooden cockpit shell with magnet-backed instrument facsimiles and seven rear projection screens to simulate the cockpit displays and many of the controls, The throttle and stick had mock switches. Seven slide projectors, positioned to project on the rear of frosted plexiglass screens, were used to display informational formats for the role-playing exercise. The specific formats projected on the individual screens were controlled by the experimenter according to the activities/tasks occurring in the flight. The role-playing scenario was non-dynamic in that the experimenter told the pilot which display to call up and which control actions to take. Upon the pilot's "activation" of a control (voice, touch, HOTAS), the experimenter manually incremented the slide projector to the next display and continued with the mission scenario.

Procedure

Prior to the first day of study participation, each pilot received an information document that contained a statement of the program objectives, a statement of the air vehicle concept, the four mission scenarios (Intercept, Counterair Sweep, Escort/CAP, and Interdiction), and descriptions of the cockpit layout, avionics moding, display and control functions, and cockpit operation logic.

Pilots participated in this study for one and a half days. Prior to role-playing the mission scenario, pilots received a briefing that covered objectives of the study, completed a three hour ground school on advanced cockpit mockup operation, and performed the scale development activity for SWAT (Reid, et. al., 1981).

1. Training Advanced Cockpit Concepts

The cockpit mockup training phase took place after the ground school and included three short mission phases: Cockpit Check, BIT Check, and Level Off. This training phase was intended to familiarize the subject with the cockpit mockup and introduce him to the test procedures and the three cockpit interface modes (voice, touch, and HOTAS).

2. Evaluation Conditions for Advanced Cockpit

The mission scenario was representative of a future advanced fighter mission. In an effort to evaluate prototype aircrew protection CBR flight gear and its effects on cockpit anthropometrics and automation concepts, two combinations of standard and prototype flight gear were examined. Each pilot role played the test scenario under each of the following experimental conditions:

(1) Standard flight gear (Helmet, Nomex gloves, anti-"G" Suit)
(2) Prototype chemical flight suit with 7 mm chemical ensemble gloves and prototype helmet shroud.

Random treatment presentation order was employed to eliminate systematic bias due to learning effects. During the testing, an experimenter directed the pilot through each mission segment and provided prompting as to what displays to bring up and which control mode (voice, touch, HOTAS) to use. The experimenter provided mission relevant information and operational descriptions to the subject pilot, and was available for questions or additional information,

3. Workload Measurement

Reflective SWAT and Projective-SWAT (Pro-SWAT) were used to estimate the workload required to perform cockpit activities during each mission segment. Prior to entering the cockpit mockup, each mission segment was described in detail and the pilot was asked to retrospectively evaluate the workload produced by imagining that he was performing each segment in the fighter cockpit in which he was most current (F-4, F-15, A-7)- Each pilot was then asked to give a Reflective SWAT rating. This data was recorded and later used as a baseline rating. After entering the cockpit mockup and role-playing a mission segment, the pilots were asked to give a Pro-SWAT rating and projectively evaluate the workload produced by performing the segment in the proposed cockpit. These Reflective and Pro-SWAT ratings formed the basis for cockpit workload comparisons.

4. Questionnaire and Structured Debrief

At the completion of two repetitions of the role-playing mission scenario and SWAT data collection, pilots responded to an extensive questionnaire. The questionnaire requested ratings and solicited comments on all display formats regarding clarity, size, information content, and other features. Control types were similarly evaluated. The questionnaire also requested subject comments on the CBR aircrew ensemble's effect on control manipulation, visibility, and pilot movement,

In a post questionnaire debriefing, pilots were interviewed in an effort to draw out additional comments and qualifications for earlier questionnaire responses, Audio recordings were made and reviewed for pertinent information.

RESULTS

The average SWAT workload scores (Reflective and Projective) obtained across all pilots for each of the three conditions (i.e., baseline fighter, mockup with CBR gear, and mockup without CBR gear) are shown in Figure 1.

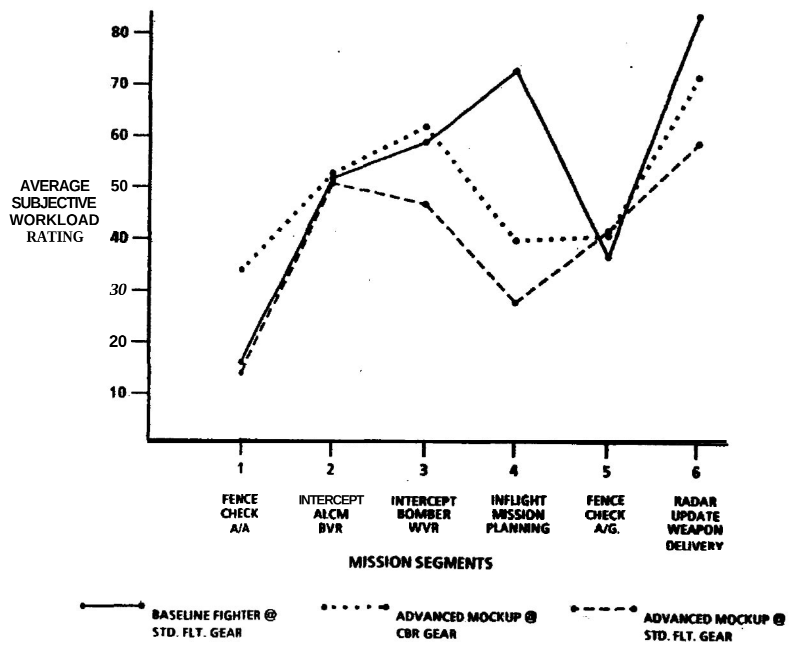

Figure 1. Average Workload Ratings Across Mission Segments

This figure shows that the pilot workload was generally the lowest while wearing the standard flight gear. For mission segments one, two, three, and five, the baseline pilot workload was lower than the advanced concept mockup workload when the pilot was wearing the CBR gear. This was not true for segments four and six, in which the advanced concept mockup workload in standard and CBR gear were both lower than the baseline pilot's workload. To determine where these differences were statistically significant (P less than or equal to 0.10), a sign test was used.

Table 1 shows the comparison between pilot workload in the baseline aircraft (reflective SWAT) and pilot workload wearing standard flight gear in the mockup (Pro-SWAT). Since the basic difference between workload in the baseline aircraft and in the mockup was thought to be largely due to the higher

TABLE 1. COMPARISON OF PERCEIVED WORKLOAD: SWAT VALUES BETWEEN BASELINE AIRCRAFT AND MOCKUP

SEGMENT NUMBER	NUMBER OF PILOTS INDICATING LOWER WORKLOAD FOR BASELINE	NUMBER OF PILOTS INDICATING LOWER WORKLOAD FOR MOCKUP	PROBABILITY OF INDICATED RESPONSES ASSUMING EQUAL WORKLOAD[a]
1	4	4	0.6367
2	3	7	0.1719
3	3	7	0.1719
4	0	12	0.0002***
5	7	4	0.8867
6	0	9	0.0020***

[a] These probabilities are the binomial tail distribution probabilities under H_0 = no difference between configurations ($p=\frac{1}{2}$), versus the alternative hypothesis, H_A = wdrkload is significantly lower in the mockup (p less than $\frac{1}{2}$).

* Significant at P less than 0.10
** Significant at P less than 0.05
*** Significant at P less than 0.01

level of automation in the advanced concept mockup, it was assumed that the workload in the mockup using standard gear should be significantly lower. The null hypothesis, if not rejected, would indicate the number of pilots believing the workload was lower for the baseline would be roughly equal to the number of pilots believing the workload was lower for the mockup. Using the binomial distribution to calculate the probability under the null hypothesis, it was determined that there is significantly (P less than or equal to 0.01) lower workload for the advanced concept mockup in the standard gear when compared to the baseline for segment four and six, but not significantly different workload for the other segments.

Table 2 shows the pilot workload comparison between flying the mockup while wearing the standard flight gear and flying the mockup while wearing CBR gear. Since the CBR gear would hinder pilot activity, the null hypothesis assumes that there was no difference between the two conditions and the alternative hypothesis assumes that the workload would be significantly lower flying the mockup while wearing the standard flight gear. When the binomial probabilities were computed, it was determined that workload was significantly lower (P less than or equal to .10) for flying the mockup while wearing standard flight gear in segments one, three, four, and six, but not for segments two and five.

Table 3 shows the workload comparison between flying the baseline aircraft and flying the mockup using CBR gear. It was not clear whether the influence of cockpit automation (expected to reduce workload) in the advanced mockup would entirely compensate for the use of CBR gear (expected to increase workload). Thus, when testing for differences, it was assumed that either there was no difference (null hypothesis) or that a difference exists (a two sided alternative hypothesis). When the binomial probabilities are computed, it is evident that the only significant difference in workload between the baseline and the mockup, while wearing CBR gear, is in segment four. Here, the advanced cockpit concept role-played using CBR gear has a significantly lower workload.

TABLE 2. COMPARISON OF PERCEIVED WORKLOAD: SWAT VALUES IN MOCKUP WITH AND WITHOUT CBR GEAR

SEGMENT NUMBER	NUMBER OF PILOTS INDICATING LOWER WORKLOAD FOR MOCKUP WITHOUT CBR GEAR	NUMBER OF PILOTS INDICATING LOWER WORKLOAD FOR MOCKUP WITH CBR GEAR	PROBABILITY OF INDICATED RESPONSES ASSUMING EQUAL WORKLOAD[a]
1	6	1	0.0625*
2	6	2	0.1445
3	7	2	0.0898*
4	6	0	0.0156**
5	5	6	0.7256
6	6	1	0.0625*

[a] These probabilities are the binomial tail distribution probabilities under H_0 = no difference between the workload in the mockup without CBR gear compared to the workload in the mockup with CBR gear, versus the alternative hypothesis H_A = the workload is significantly higher in the mockup with CBR gear.

* Significant at P less than 0.10
** Significant at P less than 0.05
*** significant at P less than 0.01

TABLE 3. COMPARISON OF PERCEIVED WORKLOAD: SWAT VALUES BETWEEN BASELINE AIRCRAFT AND MOCKUP WITH CBR GEAR

SEGMENT NUMBER	NUMBER OF PILOTS INDICATING LOWER WORKLOAD FOR BASELINE	NUMBER OF PILOTS INDICATING LOWER WORKLOAD FOR MOCKUP WITH CBR GEAR	PROBABILITY OF INDICATED RESPONSES ASSUMING EQUAL WORKLOAD[a]
1	4	3	1.0000
2	8	3	0.2266
3	6	3	0.2539
4	2	10	0.0386**
5	6	4	0.7539
6	2	8	0.1093

[a] These probabilities are the binomial tail distribution probabilities under H_0 = no difference between the workload in the baseline aircraft compared to the mockup with CBR gear, versus the two-sided alternative hypothesis H_A = workload is significantly higher in one configuration that in the other configuration.

* Significant at P less than 0.10
** Significant at P less than 0.05
*** Significant at P less than 0.01

DISCUSSION

In mission segments where increased cockpit automation and improvements in controls, displays, and avionics were expected to reduce cockpit tasking to the greatest degree, a significant reduction was found in the perceived pilot workload (segments four and six) compared to baseline fighters of today. When comparing standard flight gear with CBR protective gear in the mockup cockpit, it was found that the CBR gear significantly increased the workload in segments one, three, four, and six. Only in mission segment four, where cockpit tasking would require that a large number of mathematical calculations and mission planning activities be accomplished, was there still a significant reduction in perceived workload over the baseline, even when the CBR gear was worn.

Results of this study suggest that the design of an advanced fighter cockpit that requires its operators to wear the additional CBR protective equipment will need to consider the effects of this equipment from the earliest stages of its development. The improvements to cockpit workload brought about by cockpit automation, control interoperability, electronic displays, and advancements in avionics may be somewhat negated by the adverse effects of physically and psychologically encumbering the aircrew with CBR gear,

In the selection of mission segments for this study, the emphasis was to present a cross-section of fighter mission cockpit tasks. The results provide some indication that the improvements in cockpit automation, displays, controls, and avionics moding were not equally powerful in reducing pilot workload under all mission segments and conditions, These results could be explained in terms of the operational realities of different segments in a fighter mission. Some mission segments, like segment four, are inherently more demanding on the information processing capabilities of the pilot than others, In subsequent studies, where the benefits of advances and improvements in avionics, cockpit automation, and numerous other areas are to be evaluated against present systems, the selection of mission segments should be certain to include those areas where they could potentially make the'greatest impact,

Portions of the study designed for the structured role-playing activity were based upon the recommendations of Arbak et. al (1984) and Greene et. al, (1981), in that the interviewer and the subject pilot were both experienced and familiar with the dynamic environment of the fighter cockpit. Also, the role-playing activity itself contained detailed and scripted cockpit tasking items and was set in a mission scenario and cockpit mockup that enhanced the realism required for accurate subjective evaluation, These design features and the results they foster, if taken in the context of first cut subjective data for predesign concept evaluations, can provide useful tools for early evaluation of cockpit systems.

As in all simulation studies, the validity of subjectively based measures of workload are sensitive to the realism under which the data collection activities take place, Role-playing activities in mockup cockpits require the pilot to do a great deal of mental imaging. This activity can only best be accomplished when as many elements as possible of their experimental activities 'are operationally realistic, This means that pilots should be selected from the same category and experience level as that anticipated for the concept aircraft's pilot. Also, experimenters and their materials must be as true to the mission scenario as is reasonable, This study, by addressing and implementing these controls is believed to have illustrated that the use of role-playing activity in a mockup is a valid step in the concept evaluation process.

In this study, support was found for the sensitivity of the SWAT techniques (reflective and projective) to the role-playing activities. Significant results were found when three different conditions were compared in a role-playing exercise, This experiment provided another unique opportunity for the SWAT techniques to be used in the collection of operator-based subjective workload data. Recordings of pilot's comments as they made their ratings in addition to the scaling properties of the SWAT techniques, provide additional information that the scientist may employ to systematically examine the complex cockpit environment. Subsequent studies might combine the techniques used here with improvements in the areas of simulation dynamics and subject cueing to increase the realism of the simulated cockpit environment.

REFERENCES

Arbak, C. J., Shew, R. L., and Simons, J. C. (1984), The Use of Reflective SWAT for Workload Assessment. Proceedings of the Human Factors Society Annual Meeting.

Greene, J. H., Arbak, C. J., Courtright, J. F., and O'Donnell, R. D., Ground Attack Tactics Survey (GATS), Wright-Patterson Air Force Base, Ohio: Air Force Aerospace Medical Research Laboratory Technical Report, AFAMRL-TR-81-68, June 1981.

Reid, G. B., Shingledecker, C.A., and Eggemeier, F. T. (1981), Application of Conjoint Measurement To Workload Scale Development. Proceedings of the 1981 Human Factors Society Annual Meeting, 522-526.

Reid, G. B., Shingledecker, C. A., Hockenberger, R. L., and Quinn, T. J. (1984), A Projective Application of the Subjective Workload Assessment Technique, Proceedings of the IEEE 1984 National Aerospace and Electronics Conference.